Journal of Computational Physics **171,** 616–631 (2001)
doi:10.1006/jcph.2001.6799, available online at http://www.idealibrary.com on IDEAL®

Adjoints of Nonoscillatory Advection Schemes

John Thuburn* and Thomas W. N. Haine†

Department of Meteorology, University of Reading, PO Box 243, Earley Gate, Reading, RG6 6BB, United Kingdom; and †Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland
E-mail: swsthubn@met.rdg.ac.uk; Thomas.Haine@jhu.edu

Received July 17, 2000; revised March 3, 2001

Nonoscillatory advection schemes contain switches, so that the derivative of the numerical solution at any time step with respect to that at the previous time step may be discontinuous. In consequence, sensitivities calculated using the adjoint of the numerical scheme may be discontinuous or ambiguous. This discontinuity is not a property of the continuous advection equation; it is an artefact of the numerical schemes used to solve it. The problem is demonstrated in some simple one-dimensional test cases. We derive a result showing that there is no possibility of smoothing the switches in nonoscillatory advection schemes to remove the discontinuities while retaining an obvious and desirable scaling property. We discuss some alternative approaches to deriving the adjoint schemes needed for sensitivity calculations. © 2001 Academic Press

Key Words: advection; adjoint; nonoscillatory.

1. INTRODUCTION

There are many problems in atmospheric and oceanic science where calculation of sensitivities is required. These include (i) variational data assimilation, [e.g., 8, 25, 33]; (ii) parameter estimation or retrieval, [e.g., 12, 34]; (iii) understanding physical mechanisms behind phenomena such as lee cyclogenesis [28], extratropical cyclones [13], blocking weather patterns [22], and El Niño [19]; and (iv) quantifying the stability of realistic atmospheric flows [7], with application to generating perturbed initial states for ensemble weather forecasts [9] and for the possibility of using small pilotless airplanes to adaptively target observations for initializing weather forecasts [20].

For any given mathematical or numerical model, sensitivities can often be computed efficiently using the adjoint of the corresponding tangent linear model. For example, see [6] for an introduction to adjoints and their uses. In all of the applications mentioned above, adjoints are used for computing sensitivities because their computational cost is comparable to that of the original model, making them vastly more efficient than other methods.

0021-9991/01 $35.00

Because both the tangent linear model and its adjoint are linear, one crucial factor limiting the applicability of adjoints is nonlinearity in the original mathematical or numerical model. For example, in four-dimensional variational data assimilation, which attempts to minimize the mismatch between weather observations and the state of a weather forecast model over a period of time as well as a region of space, the nonlinearity of the atmospheric flow limits the assimilation period to about 1 day at most. For similar reasons, adjoints are of limited use for quantifying climate sensitivity [15]. The nonlinearity problem is particularly acute when the mathematical or numerical model contains discontinuities or switches. For example, the representation of cumulus clouds in a weather forecast model is usually triggered by a certain threshold value of the atmospheric stability. The most extreme such form of nonlinearity, where fields at time step $n+1$ depend discontinuously on fields at time step n, can lead to unbounded values for sensitivities. A milder form of nonlinearity, in which derivatives of fields at step $n+1$ with respect to fields at step n are discontinuous, leads to discontinuous or ambiguous values for sensitivities [31]. Of course, such unbounded or discontinuous sensitivites might reflect properties of the original physical system (for example, the sensitivity of cloud liquid water concentration to temperature $\partial c_l/\partial T$ is discontinuous at the temperature at which condensation begins); then, arguably, it may be desirable to capture them in the adjoint calculation. On the other hand, extreme nonlinearities such as switches might be introduced in a mathematical or numerical model without being present in the original physical system. Then any such large or discontinuous sensitivitives indicated by an adjoint calculation would be entirely spurious artefacts of the mathematical or numerical model and unrelated to the original physical system. It is this last possibility that is the subject of this paper, in the context of nonoscillatory advection schemes.

Godunov's theorem [11] says that any linear, monotone advection scheme is at most first-order accurate. First-order schemes, however, are regarded as too diffusive for many applications. Therefore, if we require a scheme better than first-order accurate while avoiding spurious oscillations, then we must use a nonlinear advection scheme, even though the advection problem itself is linear in the advected variable when the advecting flow is given. There are many examples of nonlinear nonoscillatory advection schemes in the literature, including schemes based on flux limiters or slope limiters (e.g., see [5, 18] for an introduction), flux-corrected transport (FCT) [4, 32], or semi-Lagrangian schemes with nonlinear interpolation to ensure preservation of monotonicity [1, 30].

This nonlinearity of nonoscillatory advection schemes raises a number of issues regarding the properties of numerical advection and how they relate to those of real advection, e.g., [26]. In this paper we investigate how the nonlinearity of nonoscillatory advection schemes affects the calculation of sensitivities, particularly when using adjoints.

In Section 2 below, we use simple examples to illustrate that the nonlinearity of some commonly used kinds of nonoscillatory advection schemes is, indeed, strong enough to lead to ambiguous sensitivities in adjoint calculations. It might be hoped that it would be possible to construct nonoscillatory advection schemes whose adjoints are well-behaved (in a sense to be made precise below). In Section 3, we show that any scheme having an obvious and desirable scaling property cannot have a well-behaved adjoint unless the scheme is fully linear. But then, by Godunov's theorem, the scheme cannot be both nonoscillatory and better than first-order accurate. This result motivates us in Section 4.1 to consider alternative strategies for constructing adjoints of advection problems, and to consider some issues that arise. All of this discussion is relevant even if no perturbations to the advecting velocity are considered. Further issues that can arise when perturbations to the advecting velocity are considered in Section 4.2.

2. EXAMPLES OF PROBLEMS WITH ADJOINTS OF NONOSCILLATORY ADVECTION SCHEMES

In this section, we use simple examples to show that the nonlinearity of typical nonoscillatory advection schemes does, indeed, lead to ambiguous results from sensitivity calculations; that is, the calculated sensitivities depend on exactly how the calculation is implemented and on whether the calculation is carried out through multiple perturbed forward integrations or using an adjoint.

For our test case, we use a one-dimensional periodic grid of 20 equally spaced points. A control forward integration of the linear advection equation

$$\frac{\partial}{\partial t}q(x,t) + u\frac{\partial}{\partial x}q(x,t) = 0, \tag{1}$$

with u constant is carried out, using one of the advection schemes discussed below. To begin with, the initial profile is taken to be a delta function, since this illustrates the problem most clearly. The initial profile is advected toward the right with a constant Courant number $u\Delta t/\Delta x$ of 0.5 for 20 steps. Let q_i^n be the advected quantity at the ith grid point after n steps. The values q_i^n are saved at every time step of the control integration; these values provide the "trajectory" in phase space about which the scheme is linearized to provide the tangent linear model and its adjoint. We then ask what is the sensitivity

$$G_j^0 = \frac{\partial J}{\partial q_j^0}$$

of a certain functional J of the final state to changes in the initial data q_j^0? For illustration we take a simple case in which J is just the value of q at the 15th grid point at the final time $J = q_{15}^{20}$.

We calculated the sensitivity in two ways. The first way (the "multirun method") used multiple perturbed forward integrations. For each grid point j, the initial value q_j^0 is perturbed to $q_j^0 + \varepsilon$ ($\varepsilon \neq 0$) and the integration rerun, yielding a modified final state with q_{15}^{20} replaced by $q_{15}^{20} + \Delta$, say. The sensitivity is then estimated as $G_j^0 \approx \Delta/\varepsilon$. If the advection scheme were linear, then this estimate for G_j^0 would be exact and would be independent of the value of ε. It would also be independent of the initial data used in the control integration. For a nonlinear scheme we might hope that G_j^0 would exist and that $\Delta/\varepsilon \to G_j^0$ as $\varepsilon \to 0$. We therefore use a small value $\varepsilon = 0.001$ (that is, small compared to the range of q values in the profile) but check for convergence by rerunning with $\varepsilon = -0.001$.

The second way of estimating the sensitivity uses the adjoint of the tangent linear model. The values of q at step n are functions of those at step $n-1$, and if those functions are differentiable, then the Jacobian

$$\frac{\partial q_i^n}{\partial q_j^{n-1}}$$

exists. Letting

$$G_j^n = \frac{\partial J}{\partial q_j^n},$$

we can then evolve the sensitivity backwards in time using

$$G_j^{n-1} = \sum_i G_i^n \frac{\partial q_i^n}{\partial q_j^{n-1}}, \tag{2}$$

that is, by multiplying the vector $\mathbf{G}^n$ by the transpose of the Jacobian matrix (e.g., [6]). If the original advection scheme is linear, then the results will be identical to those obtained with the multirun method. If, in addition, the advecting wind is constant, then it may easily be verified that (2) amounts to advecting the sensitivity backwards in time using the original advection scheme.

Figure 1 shows the results of this test case using the linear, third-order QUICKEST scheme [16]. Panels (a) and (b) show the initial and final profiles for the control integration. Panel (c) shows the sensitivity at the final time G_j^{20}. It is given by $G_{15}^{20} = 1$, $G_j^{20} = 0$ for $j \neq 15$. Panels (d) and (e) show the sensitivities to initial conditions G_j^0 calculated using the adjoint method and the multirun method, respectively. As expected, because the advection scheme is linear, the two methods yield the same sensitivities, and both agree with advecting the final sensitivity backwards using the QUICKEST scheme (compare with the mirror image of panel (b)). The sensitivity calculation is clearly well-behaved in this case.

Figure 2 shows the sensitivities obtained when the QUICKEST scheme is made nonoscillatory by adding the Universal Limiter [17]. Now the sensitivity calculated by the multirun method is no longer independent of the value of ε. In particular, the values for $\varepsilon = 0.001$ (solid curve) are quite different from those for $\varepsilon = -0.001$ (dotted curve). As $|\varepsilon|$ is decreased further, the sensitivities do not converge. Furthermore, the values obtained from the adjoint calculation are different again. The sensitivity calculation is clearly not well-behaved in this case.

This bad behavior arises because the Jacobian

$$\frac{\partial q_i^n}{\partial q_j^{n-1}}$$

does not depend continuously on the q^{n-1}'s. A simple example will illustrate the essence of the problem without getting into the details of the advection scheme. QUICKEST with the Universal Limiter, and indeed almost all other nonoscillatory advection schemes, involve the computation of the minimum (or maximum) of certain sets of numbers, or some equivalent calculation. Consider the function $f = \min(a, b)$. Then for $a < b$

$$\left(\frac{\partial f}{\partial a}, \frac{\partial f}{\partial b}\right) = (1, 0), \tag{3}$$

while for $a > b$

$$\left(\frac{\partial f}{\partial a}, \frac{\partial f}{\partial b}\right) = (0, 1). \tag{4}$$

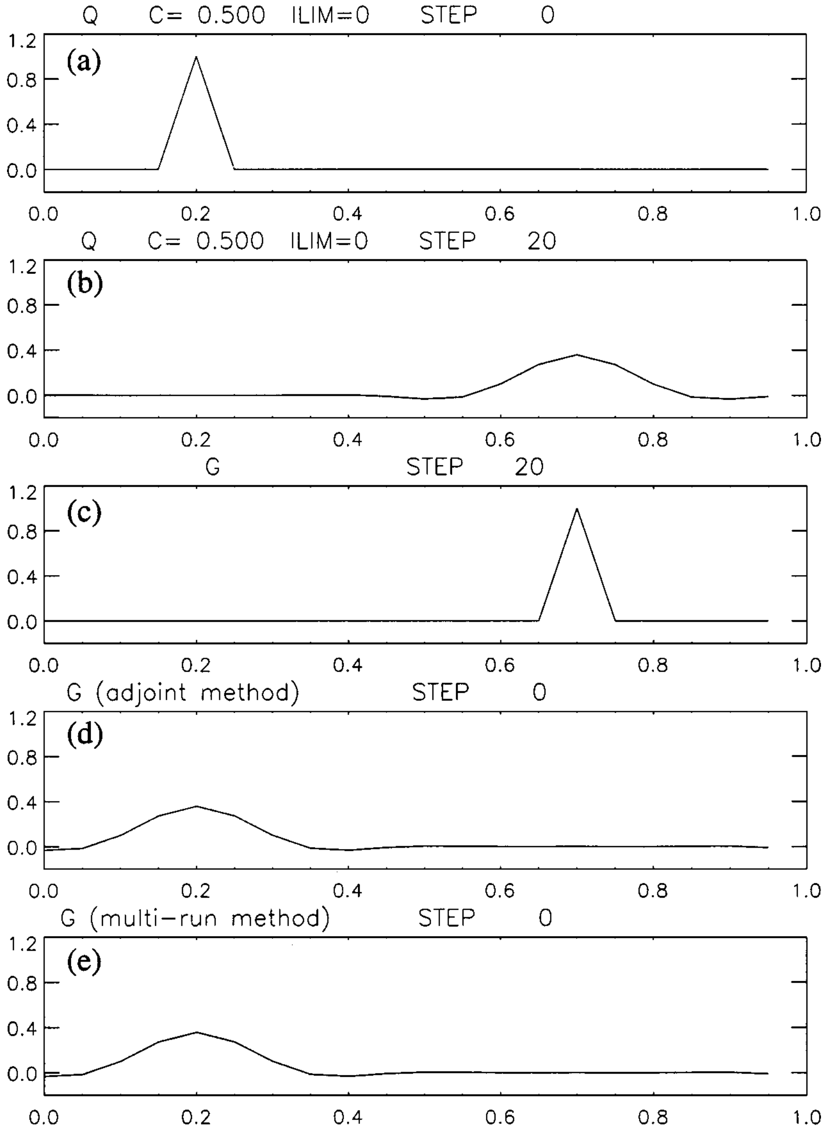

FIG. 1. Test case results for the linear QUICKEST scheme. (a) Initial condition for the control forward integration. (b) Final state for the control forward integration. (c) Sensitivity at the final time. (d) Sensitivity at the initial time computed using the adjoint. (e) Sensitivity at the initial time computed using the multirun method.

However, the partial derivatives are discontinuous at $a = b$. This explains why the multirun sensitivity calculations yield such different answers for positive and negative ε: if, at any point in the control forward integration, the situation analogous to $a = b$ (or a sufficiently close to b) occurs, then positive and negative values of ε will put the solution into two different regimes with (3) applying in one regime and (4) in the other.

This example also shows that, when a situation analogous to $a = b$ occurs,

$$\frac{\partial q_i^n}{\partial q_j^{n-1}}$$

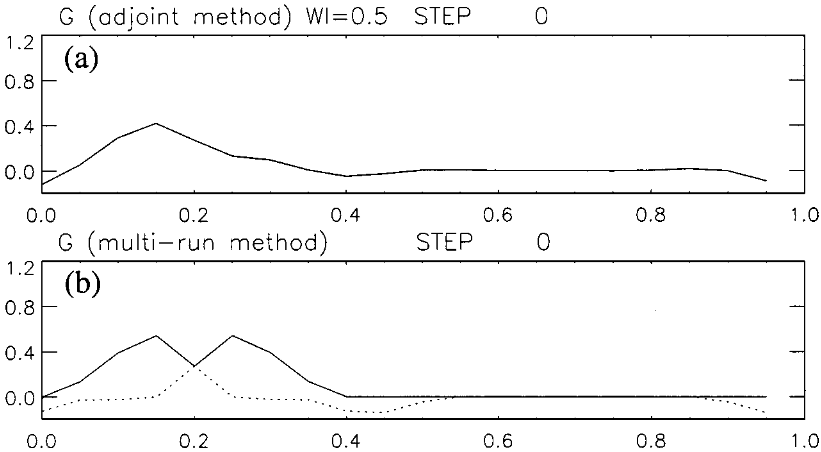

FIG. 2. Test case results for the QUICKEST scheme with the universal limiter. (a) Sensitivity at the initial time computed using the adjoint. (b) Sensitivity at the initial time computed using the multirun method; solid curve for $\varepsilon = 0.001$, dashed curve for $\varepsilon = -0.001$.

no longer exists so that, strictly, the tangent linear model and its adjoint also no longer exist. In order to obtain a sensitivity estimate from the adjoint calculation, some values for

$$\left(\frac{\partial f}{\partial a}, \frac{\partial f}{\partial b}\right)$$

must be arbitrarily specified when $a = b$. Two possibilities are given by (3) and (4). A third possibility is to take the mean of (3) and (4):

$$\left(\frac{\partial f}{\partial a}, \frac{\partial f}{\partial b}\right) = \left(\frac{1}{2}, \frac{1}{2}\right). \tag{5}$$

(There are many other equally defensible possibilities.) In fact, the result shown in Fig. 2a used the analog of (5) in the adjoint calculation. The other two possibilities, analogues of (3) and (4), give sensitivities that are different yet again (Fig. 3).

As we shall see in the next section, for most nonoscillatory advection schemes the problem with discontinuous

$$\frac{\partial q_i^n}{\partial q_j^{n-1}}$$

is most acute at places where the profile q^{n-1} is flat. This is analogous to the $a = b$ case in the simple example discussed above. Then there are q_j^{n-1} values that are on the verge of becoming extrema, and the flux limiter or monotonicity fixer in the scheme is on the verge of switching on. In view of this, the test case discussed so far is clearly quite a severe test because the control forward integration q profile is flat over many grid points. Nevertheless, this test has practical relevance because extensive areas of flat (in fact zero) values occur, for example, in the layer thickness field of isopycnal-coordinate ocean models where the layers outcrop, and a nonoscillatory advection scheme (or at least a sign preserving scheme) is essential to ensure that negative layer thicknesses are not created [3].

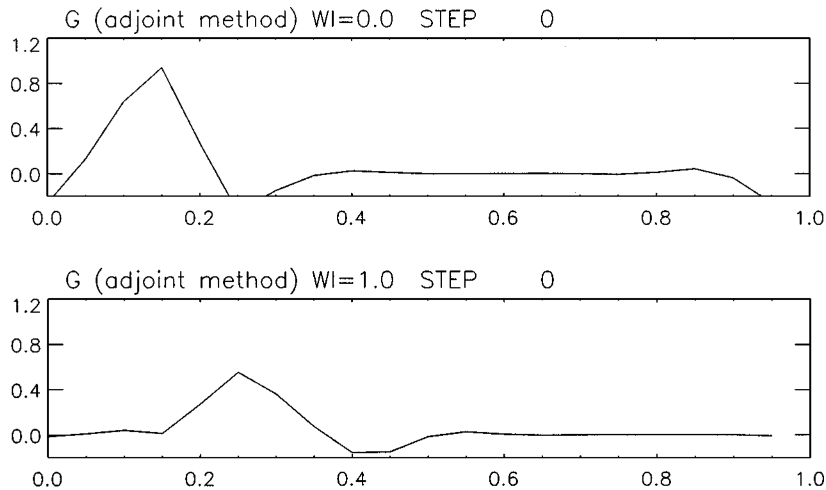

FIG. 3. Test case results for the QUICKEST scheme with the universal limiter. Panels show sensitivity at the initial time computed using two alternative approximations, analogous to (3) and (4), for the adjoint at the point where $\frac{\partial q_i^n}{\partial q_j^{n-1}}$ is discontinuous. Compare Fig. 2a.

Figures 4 to 7 show results from a less severe test case for a variety of nonoscillatory advection schemes. In this case, the initial state for the control forward integration is a wavenumber 1 sine wave, so that the strong nonlinearity associated with the flux limiter or monotonicity fixer should only be important near the maximum and minimum of the sine wave. For this test case, the scheme already discussed above (QUICKEST plus the Universal Limiter) is not very badly behaved, though there are noticeable differences between the sensitivities obtained by the adjoint method and the multirun method, and among the three versions of the adjoint method using analogues of (3), (4), and (5) above (only the case corresponding to (5) is shown). However, the sensitivity calculations using the other

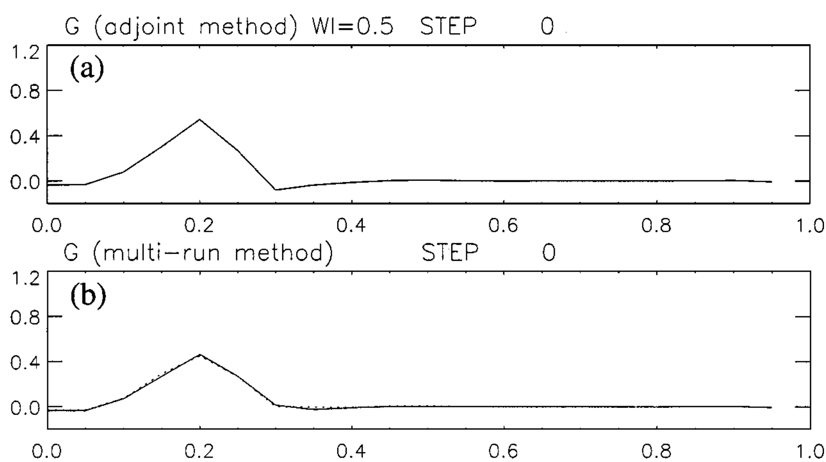

FIG. 4. Test case results for the QUICKEST scheme with the universal limiter. The initial data for the control forward integration is a wavenumber 1 sine wave. (a) Sensitivity at the initial time computed using the adjoint. (b) Sensitivity at the initial time computed using the multirun method; solid curve for $\varepsilon = 0.001$, dashed curve for $\varepsilon = -0.001$.

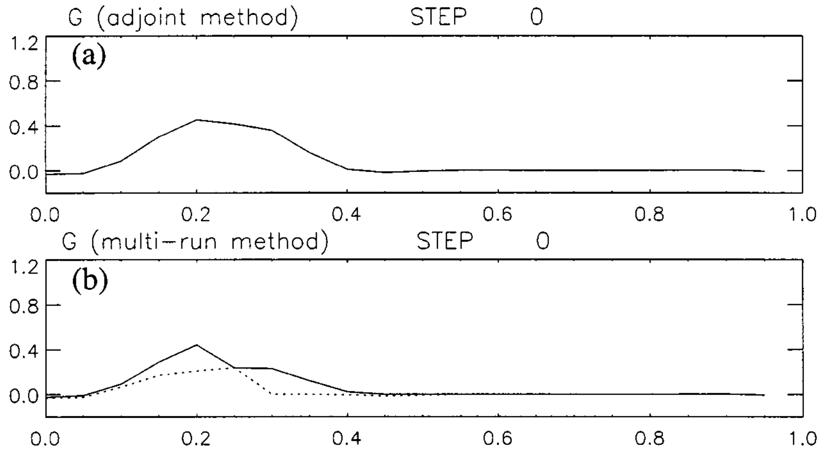

FIG. 5. As in Fig. 4, but for the QUICKEST scheme made nonoscillatory using the flux corrected transport algorithm [4, 32].

three advection schemes, which are typical of those in widespread use, are quite badly behaved.

3. INVARIANCE UNDER RESCALING WITHOUT FULL LINEARITY IMPLIES A BADLY BEHAVED ADJOINT

It might be wondered whether there is some way of constructing an advection scheme that is well-behaved in sensitivity calculations without losing other desirable properties, for example, by modifying an existing scheme so that the switches act more smoothly. In this section we prove a result implying that this goal cannot be achieved.

Let **A** be the operator corresponding to an advection scheme that maps $\mathbf{q}^{n-1}$ into $\mathbf{q}^n$:

$$q_i^n = A_i(\mathbf{q}^{n-1}). \tag{6}$$

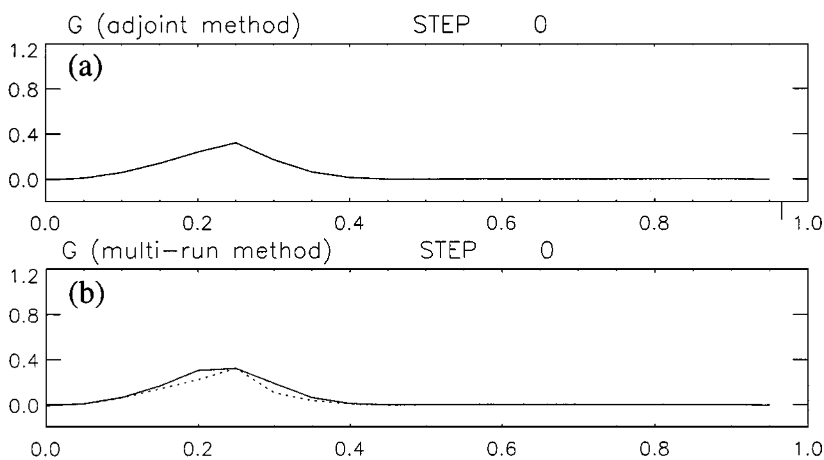

FIG. 6. As in Fig. 4, but for a total variation diminishing (TVD) scheme using the van Leer limiter [27].

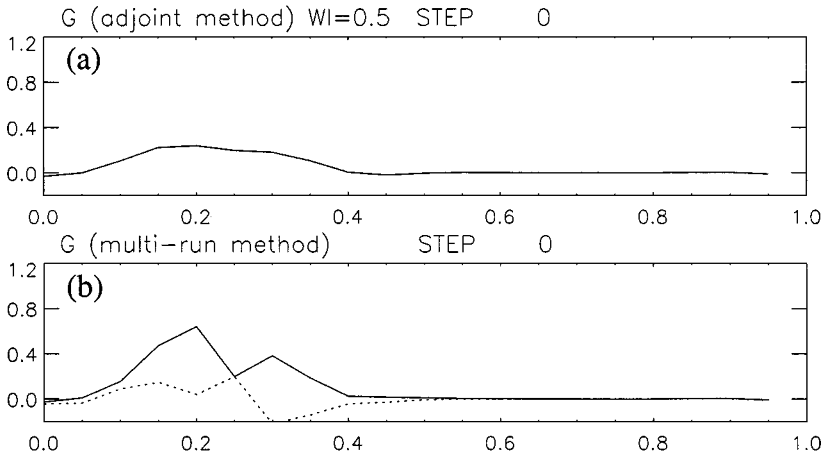

FIG. 7. As in Fig. 4, but for a semi-Lagrangian scheme using cubic Lagrange interpolation with a simple monotonicity fixer in which the interpolated value is forbidden to lie outside the range of the two nearest gridpoint values (e.g., [1]).

Many advection schemes, whether linear or not, satisfy the scaling property

$$A_j(\alpha\mathbf{q}) = \alpha A_j(\mathbf{q}) \tag{7}$$

for all profiles $\mathbf{q}$ and for any constant α. Indeed, many satisfy the stronger relation

$$A_i(\alpha\mathbf{q} + \beta\mathbf{c}) = \alpha A_i(\mathbf{q}) + \beta \tag{8}$$

for all profiles $\mathbf{q}$ and for any constants α and β, and where $\mathbf{c}$ is the constant unit profile $c_i = 1\ \forall i$. The scaling property (7) implies that rescaling the initial conditions before advecting gives the same result as rescaling by the same factor after advecting. It implies that we obtain the same physical result irrespective of what units $\mathbf{q}$ is expressed in. The stronger property (8) implies, in addition, that adding a constant to the initial conditions before advecting gives the same result as adding the same constant after advecting. Less obviously, property (8) implies that two advected quantities $\mathbf{q}_1$ and $\mathbf{q}_2$ that are initially related by a straight line functional relation

$$\mathbf{q}_2 = \alpha\mathbf{q}_1 + \beta\mathbf{c} \tag{9}$$

retain this functional relation under advection. Compact functional relations, often close to straight lines, between mixing ratios of long-lived chemicals give valuable information about chemistry, transport, and mixing in the stratosphere, for example, and it is crucial for numerical models to be able to capture those functional relations without distorting them [e.g., 26]. Finally, properties (8) and (7) are implied by full linearity, though the converse is not true. The continuous advection equation is linear in the advected quantity when the advecting flow is given, and therefore satisfies continuous analogues of (7) and (8), (where $\mathbf{A}$ is then interpreted as the operator that advects the field $\mathbf{q}$ for a finite time). For all of these reasons, the scaling properties (7) and (8) are arguably very desirable for an advection scheme, given that full linearity cannot be achieved for a nonoscillatory scheme that is better than first-order accurate. Many widely used advection schemes, and all of those discussed in Section 2 above, have property (8). Schemes that are sign preserving but not fully nonoscillatory cannot have property (8), but many widely used sign-preserving schemes [e.g., 5, 24] have property (7).

A numerical scheme of the form (6) will be well-behaved in sensitivity calculations, including adjoint calculations, if and only if the Jacobian

$$\frac{\partial A_i}{\partial q_j}$$

exists and is a continuous function of its data. Continuity of the Jacobian is also a necessary condition for the "correctness" of any tangent linear model derived from the scheme (6) [21]. It is clear that continuity of the Jacobian is indeed the property that determines whether or not sensitivity calculations give ambiguous or even unbounded results, since

$$\frac{\partial J}{\partial q_j^0} = \frac{\partial J}{\partial q_{j_n}^n} \frac{\partial A_{j_n}(\mathbf{q}^{n-1})}{\partial q_{j_{n-1}}^{n-1}} \frac{\partial A_{j_{n-1}}(\mathbf{q}^{n-2})}{\partial q_{j_{n-2}}^{n-2}} \frac{\partial A_{j_{n-2}}(\mathbf{q}^{n-3})}{\partial q_{j_{n-3}}^{n-3}} \cdots \frac{\partial A_{j_1}(\mathbf{q}^0)}{\partial q_j^0}. \tag{10}$$

(Here there is implied summation over each of the dummy subscripts $j_1, j_2, \ldots, j_n$.)

We will now show that a scheme that has both property (7) and a continuous Jacobian must be linear. First note, from the function-of-a-function rule for differentiation, that for any α

$$\left.\frac{\partial A_k(\alpha\mathbf{q})}{\partial q_j}\right|_{\text{at }\mathbf{q}=\mathbf{p}} = \alpha \left.\frac{\partial A_k(\mathbf{q})}{\partial q_j}\right|_{\text{at }\mathbf{q}=\alpha\mathbf{p}}. \tag{11}$$

Also, taking $\partial/\partial q_j$ of (7) gives

$$\left.\frac{\partial A_k(\alpha\mathbf{q})}{\partial q_j}\right|_{\text{at }\mathbf{q}=\mathbf{p}} = \alpha \left.\frac{\partial A_k(\mathbf{q})}{\partial q_j}\right|_{\text{at }\mathbf{q}=\mathbf{p}}. \tag{12}$$

Hence, since the left-hand sides of (11) and (12) are equal, the right-hand sides must also be equal

$$\left.\frac{\partial A_k(\mathbf{q})}{\partial q_j}\right|_{\text{at }\mathbf{q}=\mathbf{p}} = \left.\frac{\partial A_k(\mathbf{q})}{\partial q_j}\right|_{\text{at }\mathbf{q}=\alpha\mathbf{p}}, \tag{13}$$

provided $\alpha \neq 0$. But now the continuity of the Jacobian as $\alpha \to 0$ implies

$$\left.\frac{\partial A_k(\mathbf{q})}{\partial q_j}\right|_{\text{at }\mathbf{q}=\mathbf{p}} = \left.\frac{\partial A_k(\mathbf{q})}{\partial q_j}\right|_{\text{at }\mathbf{q}=\mathbf{0}}. \tag{14}$$

That is, the Jacobian for any profile **q** must equal the Jacobian for a zero profile. Then the partial derivatives can simply be integrated up, using $\mathbf{A}(\mathbf{0}) = \mathbf{0}$, to give

$$A_k(\mathbf{q}) = \sum_j q_j \left.\frac{\partial A_k}{\partial q_j}\right|_{\text{at }\mathbf{q}=\mathbf{0}}. \tag{15}$$

The scheme is manifestly linear, since the

$$\left.\frac{\partial A_k}{\partial q_j}\right|_{\text{at } \mathbf{q}=\mathbf{0}}$$

are constants.

The implication of this result is that an advection scheme satisfying (7) and with a continuous Jacobian cannot, by Godunov's theorem, be both nonoscillatory and better than first-order accurate. An equivalent statement of the result is that a scheme that satisfies (7) but is not linear cannot have a continuous Jacobian, and therefore cannot be well-behaved in sensitivity caculations. This means that a nonoscillatory, better than first-order accurate advection scheme (which must be nonlinear) that also satisfies (7) cannot be well-behaved in sensitivity calculations.

In deriving the above result, we did not use property (8), but only the weaker scaling property (7). Therefore, the conclusion applies not just to nonoscillatory schemes satisfying (8) but also to sign preserving schemes [e.g., 5, 24], which must also be nonlinear to be better than first-order accurate, as long as they have property (7).

It is instructive to note why the conditions of the above proof do not hold for the nonoscillatory schemes discussed in Section 2, and other nonoscillatory schemes. The scaling property implies that (13) must be satisfied. However, the Jacobian

$$\frac{\partial A_k}{\partial q_j}$$

is not continuous for a flat profile $\mathbf{q} = \mathbf{0}$ (or, in fact, for $\mathbf{q} = \beta\mathbf{c}$ for any constant β), so we cannot make the step to (14). This highlights the fact that flat profiles are the most problematic for sensitivity calculations, since arbitrarily small deviations from a flat profile lead to finite changes in

$$\frac{\partial A_k}{\partial q_j}.$$

Finally, it might be wondered whether a well-behaved nonoscillatory scheme could be obtained by abandoning the scaling property. Indeed this can be done, and an example of such a scheme is given in the Appendix. However, on top of the extra complexity of the scheme, it has several other undesirable features: (i) because the scaling property has been abandoned, the results obtained with the scheme will depend on the units that q is expressed in; (ii) an arbitrary tunable parameter must be introduced against which to measure deviations of the q profile from flatness; (iii) the scheme becomes only first-order accurate when the q profile is close to flat. Overall, the disadvantages of such a scheme are likely to outweigh the combined nonoscillatory property and good sensitivity behavior, and we would not recommend this approach.

4. DISCUSSION AND CONCLUSION

4.1. *Alternative Strategies for Building Adjoints*

The result derived in Section 3 above motivates us to consider alternative routes to constructing adjoints. In outline, there are three well-known possible routes:

(i) Discretize-linearize-adjoint;
(ii) Linearize-adjoint-discretize; and
(iii) Linearize-discretize-adjoint.

So far we have been considering route (i), in which we have a discrete numerical model and we wish to linearize and take its adjoint. Route (i) has the advantage that, given a discrete nonlinear numerical model, the linearization and construction of the adjoint can largely be automated [2, 10], greatly reducing the development effort required for complex models. A second feature of route (i) is that it leads to an exact adjoint of the original discrete nonlinear numerical model. The importance of this property is a subject of current research [e.g., 14, 23], and it is likely that the answer depends on the application. In practice, approximate adjoints have been used successfully for many applications (most of the applications mentioned in Section 1, for example, make some sort of approximation) though an exact adjoint might be crucial for some purposes. The result of Section 3 implies that if we insist on retaining a nonoscillatory advection scheme with scaling property (7) while having a well-behaved adjoint then route (i) is no longer an option.

The simplest alternative to route (i) is a variation on route (i) in which the discretizition used in constructing the adjoint differs from that in the original numerical model. For example, the flux limiter or monotonicity fixer could be removed from a nonoscillatory advection scheme for the purpose of constructing the adjoint. This retains the advantage that the linearization and adjoint stages of the construction can be automated.

Route (ii) appears to be possible using either a linear advection scheme or a nonlinear nonoscillatory advection scheme. For example, simple test cases, such as those in Section 2 in which the sensitivity is simply advected backwards using the original nonoscillatory advection scheme, yield accurate and well-behaved results. Vukićević et al. [29] reached a similar conclusion for a more realistic two-dimensional advective data assimilation problem. The use of a nonoscillatory scheme would be ruled out if linearity of the adjoint were crucial for the application and better than first-order accuracy were required (though not if property (7) were sufficient). On the other hand, the continuous advection equation implies that the sensitivity

$$\frac{\partial q(\mathbf{x}_2, t)}{\partial q(\mathbf{x}_1, 0)}$$

must be greater than or equal to zero; if it is important to capture the discrete analogue of this

$$\frac{\partial q_i^n}{\partial q_j^0} \geq 0,$$

then a nonoscillatory scheme, or at least a sign preserving scheme, must be used to advect the sensitivity backwards [29].

Route (iii) is possible only if the discretization does not introduce any new nonlinearity, as a nonoscillatory advection scheme would, for example. If the discretization does introduce nonlinearity then a second linearization stage would be needed after the discretization. In this case, the result of Section 3 would still hold, so it appears that nothing would be gained by this approach.

4.2. *Effects of Perturbed Advecting Velocity*

So far we have considered the advecting velocity to be fixed, and only the dependence of **A** on **q** has been considered. If we follow route (i) of Section 4.1 then, in the more general case, we must consider the dependence of **A** on the advecting velocity **u** too. (The values u_l may be defined at the same set of points as q_j or staggered with respect to them, depending on the grid and schemes used.) Then, in order for the tangent linear model and adjoint to be well-behaved, we require $\partial A_k/\partial q_j$ and $\partial A_k/\partial u_l$ to be continous functions of both **q** and **u**. There are two possible sources of discontinuities, related to jumps in the stencil and to switching of limiters.

Discontinuities caused by jumps in the stencil have been discussed previously, [e.g., 21]. Small changes in Courant number, usually as it crosses an integer value, can result in a different set of gridpoint q values (the stencil) being used to calculate the updated field, and hence to discontinuities in $\partial A_k/\partial q_j$ and $\partial A_k/\partial u_l$, or even in A_k itself. This potential problem is particularly relevant for semi-Lagrangian advection schemes because they are stable for large Courant numbers, so providing more opportunities for near-integer Courant numbers. Other kinds of advection schemes are usually restricted to Courant numbers between -1 and 1 but may still experience a jump in stencil and a discontinuity in $\partial A_k/\partial u_l$ as the Courant number goes through zero if the scheme is an "upwind" scheme. ($\partial A_k/\partial q_j$ should approach δ_{kj} as the Courant number approaches zero, where δ_{kj} is the Kronecker delta, and should therefore be continuous across zero.) For semi-Lagrangian schemes, it has been shown [21] that this problem can be eliminated by using interpolating functions that are continuous and have continuous derivatives across grid cell boundaries.

Discontinuities caused by switching of limiters is a distinct problem, and is the topic of this paper. Variations in either **q** or **u** might cause a limiter or monotonicity fixer to switch on or off, and both $\partial A_k/\partial q_j$ and $\partial A_k/\partial u_l$ can be discontinuous across the limiter switching point. Thus, any of the following possibilites might occur:

(i) $\partial A_k/\partial q_j$ might be discontinuous as q varies across a limiter switching point;
(ii) $\partial A_k/\partial u_l$ might be discontinuous as q varies across a limiter switching point.

For some limiters, variations in u cannot cause the limiter to switch. If variations in u can cause the limiter to switch then

(iii) $\partial A_k/\partial q_j$ might be discontinuous as u varies across a limiter switching point;
(iv) $\partial A_k/\partial u_l$ might be discontinuous as u varies across a limiter switching point.

In the preceding sections we have examined possibility (i) and shown that this kind of discontinuity is unavoidable for a nonoscillatory scheme satisfying the desirable scaling property (7). A couple of examples will illustrate that the remaining possibilities may or may not occur.

For a total variation diminishing (TVD) scheme with the van Leer [27] limiter, $\partial A_k/\partial u_l$ is in fact continuous as q varies across the limiter switching point, so possibility (ii) does not occur, even though $\partial A_k/\partial q_j$ is discontinuous. Also, variations in u do not cause the limiter to switch, so possibilities (iii) and (iv) do not occur.

For the QUICKEST scheme [16] with the universal limiter [17], on the other hand, $\partial A_k/\partial u_l$ is discontinuous as q varies across a limiter switching point, and also variations in u can cause the limiter to switch and both $\partial A_k/\partial q_j$ and $\partial A_k/\partial u_l$ will generally be discontinuous as u varies across the limiter switching point. Thus, all four possibilities (i)–(iv) occur for this scheme.

4.3. *Conclusion*

Nonoscillatory advection schemes can lead to ambiguous results in sensitivity calculations, whether those calculations are carried out via multiple perturbed forward integrations or by using the adjoint of a linearization of the scheme. Examples of the problem have been shown in a simple test case for several schemes typical of those widely used. We have shown that this sort of problem is unavoidable for nonoscillatory or sign-preserving schemes that are better than first-order accurate and therefore nonlinear, unless a certain desirable scaling property is given up. Consequently, alternative routes to constructing adjoints for advection problems must be considered, and another work [29] has shown that the linearize–adjoint–discretize route can be successful.

APPENDIX: A NONOSCILLATORY SCHEME WITH A CONTINUOUS JACOBIAN

Consider advection schemes of the form

$$q_j^{n+1} = q_j^n + c(\hat{q}_{j-1/2} - \hat{q}_{j+1/2}), \tag{16}$$

where c is the Courant number, taken to be constant and positive here for simplicity. Different choices for determining the $\hat{q}$ values yield different advection schemes. The simplest nonoscillatory advection scheme with a continuous Jacobian is the first-order upwind scheme given by

$$\hat{q}_{j+1/2} = q_j^n. \tag{17}$$

A more accurate nonoscillatory scheme with continuous Jacobian is given by

$$\hat{q}_{j+1/2} = q_j + \frac{1}{2}(1-c)\phi(r_j)\Psi(s_j)(q_{j+1} - q_j), \tag{18}$$

where q's are now understood to be evaluated at step n unless otherwise indicated. Here

$$r_j = \frac{q_j - q_{j-1}}{q_{j+1} - q_j}$$

and $\phi(r)$ is one of the well-known van Leer limiter functions [27]

$$\phi(r) = \frac{r + |r|}{1 + |r|}.$$

If the function Ψ were identically unity then this scheme would be a familiar TVD scheme (satisfying (7)), but which has discontinuities in $\partial A_k/\partial q_j$ for flat profiles and where $r_j \to 0$ or ∞. To remove these discontinuities, Ψ is defined to be a function that controls a smooth transition from a first-order upwind scheme ($\Psi = 0$) for flat profiles to the original TVD scheme ($\Psi = 1$) at large amplitude:

$$2\Psi(s) = \frac{s + |s|}{1 + |s|},$$

where

$$s_j = \frac{(q_{j+1} - q_j)(q_j - q_{j-1})}{a^2},$$

and a is a tunable parameter that defines the amplitude for the transition. It may be verified that $\partial A_k / \partial q_j$ is indeed continuous for the resulting scheme. However, the resulting scheme no longer satisfies (7).

ACKNOWLEDGMENTS

We are grateful to three anonymous reviewers for their constructive comments on an earlier version of this paper.

REFERENCES

1. R. Bermejo and A. Staniforth, The conversion of semi-lagrangian advection schemes to quasi-monotone schemes, *Mon. Weather Rev.* **120**, 2622 (1992).
2. C. Bischof, Automatic differentiation, tangent linear models, and pseudo-adjoints, in *High-Performance Computing in the Geosciences, Series C: Mathematical and Physical Sciences*, edited by F.-X. Le Dimet (Kluwer Academic, Dordrecht, Norwell, MA, 1994), pp. 59–80.
3. R. Bleck and D. Boudra, Wind-driven spin-up in eddy-resolving ocean models formulated in isopycnic and isobaric coordinates, *J. Geophys. Res.* **91C**, 7611 (1986).
4. J. P. Boris and D. L. Book, Flux Corrected Transport. I. SHASTA, a fluid transport algorithm that works, *J. Comput. Phys.* **11**, 38 (1972).
5. D. R. Durran, *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics* (Springer-Verlag, Berlin/New York, 1999).
6. R. M. Errico, What is an adjoint model? *Bull. Amer. Met. Soc.* **78**, 2577 (1997).
7. B. F. Farrel and P. J. Ioannou, Generalized stability theory. Part I: Autonomous operators, *J. Atmos. Sci.* **53**, 2025 (1996).
8. M. Fisher and D. J. Lary, Lagrangian 4-dimensional variational data assimilation of chemical-species, *Q. J. R. Meteorol. Soc.* **121**, 1681 (1995).
9. R. Gelaro, R. Buizza, T. N. Palmer, and E. Klinker, Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors, *J. Atmos. Sci.* **55**, 1012 (1998).
10. R. Giering and T. Kaminski, Recipes for adjoint code construction, *Assoc. Comput. Mach., Trans. Math. Software* **24**, 437 (1998).
11. S. K. Godunov, Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics (in Russian), *Mat. Sb.* **47**, 271 (1959).
12. A. K. Griffith, *Data Assimilation for Numerical Weather Prediction Using Control Theory*, Ph.D. Thesis (University of Reading, UK, 1997).
13. R. H. Langland, R. L. Elsberry, and R. M. Errico, Evaluation of physical processes in an idealized extratropical cyclone using adjoint sensitivity, *Q. J. R. Meteorol. Soc.* **121**, 1349 (1995).
14. A. S. Lawless and N. K. Nichols, *A Comparison of Two Methods of Deriving the Tangent Linear Model*, Technical Report 257 (UK Meteorological Office Forecasting Research, 1998).
15. D. J. Lea, M. R. Allen, and T. W. N. Haine, Sensitivity analysis of the climate of a chaotic system, *Tellus* **52A**, 523 (2000).
16. B. P. Leonard, A stable and accurate convective modelling procedure based on quadratic upstream interpolation, *Comput. Meth. Appl. Mech. Eng.* **19**, 59 (1979).
17. B. P. Leonard, The ULTIMATE conservative difference scheme applied to unsteady one-dimensional advection, *Comput. Meth. Appl. Mech. Eng.* **88**, 17 (1991).
18. R. J. LeVeque, *Numerical Methods for Conservation Laws* (Birkhäuser, Basel, 1992).

19. G. J. van Oldenburgh, G. Burgers, S. Venzke, C. Eckert, and R. Giering, Tracking down the ENSO-delayed oscillator with an adjoint OGCM, *Mon. Weather Rev.* **127**, 1477 (1999).

20. T. N. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza, Singular vectors, metrics, and adaptive observations, *J. Atmos. Sci.* **55**, 633 (1998).

21. S. Polavarapu, M. Tanguay, R. Ménard, and A. Staniforth, The tangent linear model for semi-Lagrangian schemes: linearizing the process of interpolation, *Tellus* **48A**, 74 (1996).

22. M. S. F. V. de Pondeca, A. Barcilon, and X. L. Zou, A adjoint sensitivity study of the efficacy of modal and nonmodal perturbations in causing model block onset, *J. Atmos. Sci.* **55**, 2095 (1998).

23. Z. Sirkes and E. Tziperman, Finite difference of adjoint or adjoint of finite difference? *Mon. Weather Rev.* **125**, 3373 (1997).

24. P. K. Smolarkiewicz, Comment on "A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes," *Mon. Weather Rev.* **117**, 2626 (1989).

25. J. N. Thépaut and P. Courtier, Four-dimensional data assimilation using the adjoint of a multilevel primitive equation model, *Q. J. R. Met. Soc.* **117**, 1225 (1991).

26. J. Thuburn and M. E. McIntyre, Numerical advection schemes, cross-isentropic random walks, and correlations between chemical species, *J. Geophys. Res.* **102**, 677 (1997).

27. B. Van Leer, Towards the ultimate conservative difference scheme. II: Montonicity and conservation combined in a second order scheme, *J. Comput. Phys.* **14**, 361 (1974).

28. T. Vukićević and K. Raeder, Use of an adjoint model for finding triggers for alpine lee cyclogenesis, *Mon. Weather Rev.* **123**, 800 (1995).

29. T. Vukićević, M. Steyskal, and M. Hecht, Properties of advection algorithms in the context of variational data assimilation, *Mon Weather Rev.* **129**, 1221 (2001).

30. D. L. Williamson and P. Rasch, Two-dimensional semi-Lagrangian transport with shape-preserving interpolation, *Mon. Weather Rev.* **117**, 102 (1989).

31. Q. Xu, Generalized adjoint for physical processes with parameterized discontinuities. Part I: Basic issues and heuristic examples, *J. Atmos. Sci.* **53**, 1123 (1996).

32. S. T. Zalesak, Fully multidimensional flux corrected transport algorithms for fluids, *J. Comput. Phys.* **31**, 335 (1979).

33. X. Zou, I. M. Navon, and J. G. Sela, Variational data assimilation with moist threshold processes using the NMC spectral model, *Tellus* **45A**, 370 (1993).

34. D. Zupanski, A general weak constraint applicable to operational 4DVAR data assimilation systems, *Mon. Weather Rev.* **125**, 2274 (1997).